DOCUMENT RESUME

ED 178 568 TM 009 712

AUTHOR Egelston, Richard L.
TITLE Some Interesting Characteristics of Markov Chain Transition Matrices.
INSTITUTION State Univ. of New York, Albany. Educational Research and Service Center.
PUB DATE 25 Oct 78
NOTE 31p.; Paper presented at the annual meeting of the New England Educational Research Association (9th, Ellenville, NY, October 25, 1978) ; Revised May 13, 1979

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Matrices; *Probability Theory; Simulation; Statistical Analysis
IDENTIFIERS *Markov Chain; Monte Carlo Method; Stochastic Processes

ABSTRACT

A Monte Carlo investigation of Markov chain matrices was conducted to create empirical distributions for two statistics created from the transition matrices. Curve fitting techniques developed by Karl Pearson were used to deduce if theoretical equations could be fit to the two sets of distributions. The set of distributions which describe the variance among the transition probabilities in the steady state condition was discovered to be a member of the Beta distribution family. The set of distributions for t, the statistic which describes the number of transitions required to reach the steady state condition, could not be fit to any known distribution function using the Pearson techniques. Empirical results agreed very closely with the theoretical predictions. (Three types of Markov chains--regular, absorbing, and cyclic--are discussed within the paper. Information relative to the percentage of occurence of each type of chain within the population of transition matrices is presented, along with a comparison of the occurence of each type within the samples of 5,000 and 20,000 matrices generated).
(Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard L. Egelston

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Some Interesting Characteristics Of Markov Chain Transition Matrices

by

Richard L. Egelston

Presented at the Ninth NERA Annual Convocation
Ellenville, NY, October 25, 1978
Revised May 13, 1979

CENTER FOR RESEARCH

STATE UNIVERSITY OF NEW YORK
AT ALBANY

Abstract

A Monte Carlo investigation of Markov chain transition matrices was conducted to create empirical distributions for two statistics created from the transition matrices. Curve fitting techniques developed by Karl Pearson were used to deduce if theoretical equations could be fit to the two sets of distributions. The set of distributions, for V_α , the statistic which describes the variance among the transition probabilities in the steady state condition, were discovered to be members of the Beta distribution family. The set of distributions for t , the statistic which describes the number of transitions required to reach the steady state condition, could not be fit to any known distribution function using the Pearson techniques.

Three types of Markov chains (regular, absorbing, and cyclic) are discussed within the paper. Information relative to the percentage of occurance of each type of chain within the population of transition matrices is presented, along with a comparison of the occurance of each type within the samples of 5,000 and 20,000 matrices generated. Empirical results agreed very closely with the theoretical predictions.

Some Interesting Characteristics of Markov Chain Transition Matrices

Within the behavioral sciences it has been customary to account for changes in behavior of individuals and groups through the use of linear statistical models such as analysis of variance, regression analysis, and discriminant analysis. Most of these statistical models require data which is interval in nature and normally distributed. While there are non parametric counterparts to some of the parametric models mentioned above, these models are generally unable to handle repeated measures. Few mathematical models for the analysis of categorical data exist for the behavioral scientist, and categorical data models which can be used in a predictive manner are even less common. One class of models which can be used in a predictive fashion is the Markov processes, and the most useful members of this class are the Markov chains. These models are especially useful when the same categorical observation is made upon one or more people on several occasions. Interaction analysis of group behavior is one application for Markov chains. This paper reports the results of a Monte Carlo investigation of Markov chains containing two to seven categories of classification.

Introduction

The core concept of a Markov chain is the transition matrix (P). The elements of the transition matrix are called transition probabilities which represent the probability of a random process transferring to any category conditional upon being in any category on a series of previous observations. The simplest Markov chain has transition elements (p_{ij}) which describe the probability of changing categories on two successive occasions.

In general, for k observations, $k-1$ different transition matrices may be formed. However, if the probabilities of change between successive categories are independent of when the observations are made, a single transition matrix may be applicable for the entire sequence of observations. This property distinguishes the Markov chain from the more general Markov process.

Table 1 illustrates the transition data from a possible series of observations gathered over a period of time. The fictitious data could represent the test grades given by a particular teacher over a period of one year. The probability of getting an A on one test following an A on the previous test is .5. The probability of a B following an A is .4, and the probability of a C following an A is .1. Similarly, the probabilities of getting the sequences BA, BB, and BC are .3, .6, and .1, respectively.

One constraint upon the transition matrix is that the transition probabilities for each row of the matrix must sum to 1.0. Thus, each row of the transition matrix will have $m-1$ degrees of freedom when there are m categories of classification for the object. Each row of Table 1 fulfills this property. The transition matrix will have $m(m-1)$ degrees of freedom.

Insert Table 1 about here

A transition matrix may be formed from a very large number of observations on a single individual, or from a lesser number of observations pooled over a group of individuals. Either case may contain a single string

of observations from one observation session, or from several strings of observations each formed from a single session.

A transition matrix (P) gives the probability of changing from category i to category j between successive observation. If the transition matrix is constant over all observations the probability of being in category j starting from category i after exactly 2 observations is given by $P \times P = P^2$. Similarly after exactly t observations the probability of being in category j starting from category i is given by P^t.

There are three basic types of Markov chains: regular, cyclic, and absorbing. They are classified by the value of the transition probabilities which may or may not be identical to zero or unity.

Regular chains. The regular Markov chain is characterized by the fact that all elements of some power of P will be non-zero and positive. The only situation in which a zero will not disappear is the one where the probability of going to category j is identically zero for all states. That is, the object leaves state j immediately and never returns to it.

One important characteristic of regular chains is that successive powers of the transition matrix will reduce the variation among the row elements (V_α) until there is no discernable change among the row elements from one power to the next. Table 2 illustrates this convergence property of the regular chain. $P^8 = P^9 = P^k$ to the nearest thousandth in each respective element. When the transition matrix has been powered to convergence, all rows of the matrix will be identical, and each row is called a limiting vector. The converged matrix is called A, and the elements of A represent

the proportion of observations the object will spend in each category over a "large number of observations."

The phrase "large number of observations" used above is somewhat arbitrary, and with mathematical exactness means "approaching" infinity. In the practical sense, however, convergence can be assumed if the precision of the probabilities is at the level the investigator desires. Thus, the precision level for the data represented by Table 2 is to the nearest thousandth.

Insert Table 2 about here

The number of times the transition matrix must be powered to convergence is called t. Obviously the value for t will depend upon the tolerance specified by the investigator. The tolerance used for convergence in Table 2 was .0005.

Cyclic chains. Not all transition matrices can be powered to convergence. Some transition matrices have transition probabilities equal to zero which are placed in such a fashion that the object will cycle between one subset of states and another subset of states. The top of Table 3 illustrates a cyclic chain with 2 categories of classification.

Insert Table 3 about here

Absorbing chains. Some transition matrices have unity in at least one position along the main diagonal. Once the object enters the category it cannot leave. The bottom entry in Table 3 illustrates an absorbing chain

with two categories of classification including one absorbing category.

Reducible chains. Reducible chains are those whose categories can be partitioned into chains possessing the above three types.

Method

In an attempt to study the distributions of two statistics (t and V_{ol}) generated from Markov chains containing two through seven states, sample sizes of 5,000 and 20,000 transition matrices were randomly generated. The procedure for randomly generating transition matrices was done on a row by row basis. Initially a random assignment was performed with the generation of a sequence of m locations within the row. Then a random selection of transition probabilities was performed by generating two-digit numbers, and placing them in the row locations specified by the random assignment sequence. When the first row of the transition matrix was completed, the process was repeated for the second and subsequent rows until the matrix was completed. The random number generation used (RAND) was a version of RANDU SUBROUTINE (IBM, 1966, p.47), which was modified for use on the CDC 6400.

Population size. The size of the population of transition matrices was determined in order to select a large enough random sample to generate all three types of transition matrices; regular, cyclic, absorbing. The size of the population of all chains is given by the relation by $N = W^m$ where $W = C_{k+m-2,\ m-1}$, m is the number of categories, k is the number of possible values of each transition probability, and C is the standard combinatorial symbol (Egelston, 1969; pp. 99-101). The probability of obtaining absorbing chains is given by the relation $Pa = 1/k^m + \sum_{i}^{m} 1/h^m$ and number of absorbing chains

is $Na = N*Pa$. The number of cyclic chains for most orders of metrics has not been determined at this time, but the number is likely to be very small compared with absorbing chains. For example, there is exactly one cyclic chain for transition matrices with two categories, and there are exactly two cyclic chains from non-reducible transition matrices containing three categories. In addition, the three category transition matrices contain 61206 reducible cyclic chains from a population of 1.4×10^{11} transition matrices, but this number is less than .005 percent of the total number of three-state matrices. Table 4 summarizes the population sizes.

Insert Table 4 about here

Results and Discussion

Univariate and bivariate distributions of t and V_{α} were created for each order of Markov chain containing two through seven categories. Tables of percentage, cumulative percentage, percentile ranks, and distributional statistics were created for each univariate distribution. Pearson curve-fitting techniques (1896, 1901, 1916) were applied to each distribution to determine the form of the distribution.

Distributions of t. Table 5 reports the cumulative proportion distributions for the number of powers the transition matrix must be raised until convergence has been achieved within an assumed tolerance of 0.01. Standard percentile ranks were used with the low percentiles corresponding to small values of t. It is apparent from Table 5 that the distribution was more dispersed with increasing orders of transition matrices. For example,

the P_1 to P_{99} range is 51 powers for matrices of order two, while the same range for matrices of order seven is 90 powers.

Insert Table 5 about here

It is more likely that a matrix will converge with few transitions if it is of low order. This fact is apparent from the first percentile rank. No powerings are necessary for convergence of some order two matrices, while five powerings are minimally necessary for convergence with matrices of seven states. Over ten percent of the matrices converge with one powering (two transitions) for matrices of order two, while less than one percent converge with two transitions for any other order. An inspection of Table 5 reveals that similar conclusions may be drawn for other orders of chain.

Researchers may have several criteria in mind when they select a model to analyze their data. One of these criteria may be to choose a set of states which will allow as fine a discrimination as possible to be made among subjects, and still insure convergence with a small number of observations. Table 5 may be utilized for this purpose. For example, an experimenter might plan on making twn observations on a group of subjects. If it is important that predictions of terminal state occupancy be confirmed or denied, then the investigator should be fairly certain that the chain will converge within eight powerings (nine transitions). Table 5 indicates that about eighty percent of the two-state chains and only thity percent of the seven-state chains converge by ten transitions (tolerance of 0.01). The

more vital the need to verify predictions, the fewer states would be chosen. Care would have to be taken to insure that all assumptions of the model were met, no matter how many states were used. While terminal distributions (limiting vector probabilities) may be important in some cases, predictions made for only three or four stages (P^3 or P^4) in advance, may be more useful in the dynamic society of today.

Table 6 reports percentages and cumulative percentages for t, and shows one percent of the transition matrices of order two converge with a single transition (no powerings). Thirteen percent more matrices of the same order converge with one additional transition. Only with three-state transition matrices have any converged with two transitions, and only then in rare cases. This result may have been predicted in advance, since the degree of convergence is related to the similarity of column elements. It is much more likely for the columns to be similar in the transition matrix with only two elements per column than with more elements.

Insert Table 6 about here

Distributions of V_α. Table 7 reports the relative percentage of occurrence of variations among the elements of the limiting vector V_α. All distributions are J-shaped and have the mode at the minimum value of V_α (using intervals of 0.01). The range of possible values for V_α decreases as the number of states in the chain increases. Both tables show that the 0.00 values of V_α captures a greater percentage of the transition matrices as the order of chain increases from three to seven states. Only the lowest

five values (in some cases three or four values) occur with a probability greater than 0.05 utilizing either tolerance criterion.

Insert Table 7 about here

Cumulative percentages of the V_α distributions are reported in Table 8. More than one-half of the transition matrices generated for each order of chain possess a V_α of 0.02 or less. V_α values of 0.05 or less are possessed by two-thirds of the two-state chains. This proportion increased with the other orders of chain to a high of 95 percent.

Insert Table 8 about here

Pearson (1914, pp. 66-67, 1956a, pp. 41-112, 1956b, pp. 359-375, and 1956c, pp. 529-557) provided a means by which a density function could be derived from empirical data if the data were unimodal or U-shaped. Two criteria, κ_1 and κ_2, are computed from the mean, the second, third, and fourth moments about the mean, and measures of skewness and kurtosis. Each of the Pearson curves has a unique value or range of values for κ_1 and κ_2 (Pearson, 1956b, p. 361). Table 9 illustrates the set of computed values for all statistics needed to determine the criterion and, in addition, reports the sample size and correlation between t and V_α (tolerance criterion 0.01).

Insert Table 9 about here

According to Pearson, if κ_2 is greater than unity, the distribution function is a Type VI function with the form

$$f(x) = k(x-a)^q x^{-p} \qquad (1)$$

where $a \leq x \leq \infty$ and $p \geq q - 1$. If κ_2 is negative, the distribution function is a Type I function with the form

$$f(x) = \frac{1}{B(p,q)} x^{p-1} (1-x)^{q-1} \qquad (2)$$

with $0 \leq x \leq 1$ and $p, q > 0$. However, the Type I function may be transformed to the Type VI function through the substitution $y = a/x$. The Type I curve is J-shaped if one of the parameters is between 0.0 and 1.0 (Kendall and Stuart, 1963, p. 151). Both distributions are Beta distributions. However, the empirical distributions for t (under both tolerance criteria) all had K_2 values greater than 15, indicating with apparent certainty that the distributions were of the Type VI variety.

Table 9 indicates that as the number of states in the chain increases, the distribution shifts to the right and generally exhibits less skewness (β_1) and kurtosis (β_2). The two-state and seven-state distributions were exceptions to the indicated trend.

Equating the obtained mean and variance with expected values of the mean and variance from the hypothesized Beta distributions yielded two equations for each distribution which were solved simultaneously for the parameters (p and q) of the distributions. When the equations were solved for the distributions of t, all values of p were found to be positive, and all values of q were found to be negative. Therefore,

the form of the distribution was

$$f(x) = \frac{k}{x^p\ (x-a)^q} \text{ with } k \propto \frac{a^p}{B(p,q)} \text{ and } x \geq a \text{ if } a \geq 0 \text{ or} \qquad (3)$$

$x = 0$ if $a < 0$.

But this distribution function is the Pearson Type VII function (Pearson, 1956c, p. 544) which is J-shaped. Obviously, the data obtained are not J-shaped, and a contradiction exists. Furthermore, the starting points for x are beyond the range of observations, unless a

But this distribution function was the Pearson Type VII function (Pearson, 1956c, p. 544) which is J-shaped. Obviously, the data obtained were not J-shaped, and a contradiction exists. Furthermore, the starting points for x were beyond the range of observations, unless a is taken to be negative, which would imply imaginary functions (p is not an integer). The inference to be drawn was that the obtained distribution functions were not Pearson curves in spite of the fact that the criteria conform to those stated by Pearson. Pearson himself twice found incongruities in his system of classifying curves (1956c, p. 531).

Table 10 reports the parameters for the distributions of V_α (tolerance 0.01). All values were in the allowable range, and the curve plotted from the theoretical Beta distribution corresponded very closely to the obtained data. The only major discrepamcy in the fitted curve occurred at the left-hand tail of the distribution, since the fitted curve was asymptotically infinite.

Insert Table 10 about here

Because data collected in educational research is frequently precise to only two places, an interval of 0.01 was used to create the distributions of V_{cl}. If any other interval size were used, the parameters of the fitted distribution would be different, and the curve might have a different form.

The values for the parameter p decreased, and the values for the parameter q increased as the number of states in the transition matrices increased. The change in both p and q was nearly linear from state three to state seven. In fact, the correlation between m and p was -0.98, and the correlation between m and q was 1.00. Consequently, the distribution function for the transition matrices with eight or more states would probably be predictable from the results given in Table 10.

Non-regular transition matrices. Within several samples of transition matrices, some were generated which failed to converge. Each of these matrices were identified and classified by type. Within all of the distributions only one cyclic chain was created, and only one reducible chain was generated. These low frequencies confirm the relative rarity of the two types of matrices occurring by chance.

The number of absorbing chains ranged from 41 to 99 as shown by Table 11. For the two-state distribution, 99 absorbing transition matrices were generated. This figure represented 1.98 percent of the population, exactly the percentage predicted. The percentage of absorbing chains for the other five orders of chain were within 0.4 percent of the predicted quantity of one percent.

Insert Table 11 about here

Summary

Stochastic processes are processes which develop and continue to operate over a period of time. The mathematical models of stochastic processes can be used to represent repeated observations of a categorical variable. Estimates of the transition probabilities can be generated from the observed data, and transition matrices for Markov chains can also be created for these processes. The transition matrices may then be used to predict a number of things, provided initial starting state probabilities were known.

Empirical distributions of two statistics were generated for Markov chains containing states two through seven. The first statistic, t, provides a description of how many observations are required before the process reaches a steady-state condition. The set of distributions for t were similar in shape, although the moments of the distribution functions increased with an increase in the number of states. The specific equations for the data could not be determined with Pearson curve-fitting techniques.

The second set of distributions described the variance among the elements in the steady-state condition V_{∞}. These distributions were discovered through Pearson curve-fitting techniques to be Beta distributions. Both t and V_{∞} are useful indices for describing transition matrices.

Equations for predicting the percent of regular, cyclic, and absorbing

chains in the population of transition matrices were also presented. Empirical verification of these percentages were obtained using samples of 5,000 and 20,000 randomly generated transition matrices.

Non-regular transition matrices seldom occur in the general population of transition matrices. When they do occur, they are most likely to be absorbing chains. With the exception of order two matrices, non-regular chains approximate one percent of the population. With the two-state chains, the non-regular chains approximate two percent of the population.

In this study, it was assumed that the transition probabilities were drawn from rectangular distributions within rows. This assumption may not be entirely adequate to explain phenomena of real data. Many transition matrices reported in the literature have large recurrence probabilities. A study of diagonally dominated transition matrices would be of value.

References

Egelston, R.L. An empirical investigation of some mathematical porperties of stationary Markov chain transition matrices containing two through seven states. (Doctoral dissertation, State University of New York at Buffalo, 1969.) Dissertation Abstracts International, 1970, 30, pt. 3, 3779 A (University Microfilms #69-20, 597).

International Business Machines. System/360 Scientific Subroutine Package 360A-CM-03X, Programmers Manual. Author, 1966.

Pearson, Karl. Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. Philosophical Transactions of the Royal Society of London, (1896), Series A, Vol. 187, 253-318. (Early Statistical Papers. London: Cambridge University Press, 1956. 41-112). a.

Pearson, Karl. Mathematical contributions to the theory of evolution. X. Supplement to a memoir on skew variation. Philosophical Transactions of the Royal Society of London, (1901), Series A, Vol. 197, 443-459. (Early Statistical Papers. London: Cambridge University Press, 1956. 359-375). b.

Pearson, Karl. Mathematical contributions to the theory of evolution. XIX. second supplement to a memoir on skew variation. Philosophical Transactions of the Royal Society of London, (1916), Series A, Vol. 216, 429-457. (Early Statistical Papers. London: Cambridge University Press, 1956. 529-557). c.

Table 1

Sample Transition Matrix With Three Categories

		Observation x + 1		
		Category A	Category B	Category C
Observation x	Category A	.50	.40	.10
	Category B	.30	.60	.10
	Category C	.10	.20	.70

Table 2

Sample Transition Matrix Illustrating Convergence

$$P = \begin{matrix} \\ \underline{A} \\ \underline{B} \end{matrix} \begin{matrix} \begin{matrix} \underline{A} & \underline{B} \end{matrix} \\ \begin{bmatrix} .610 & .390 \\ .250 & .750 \end{bmatrix} \end{matrix}$$

$$P^2 = \begin{matrix} \\ \underline{A} \\ \underline{B} \end{matrix} \begin{matrix} \begin{matrix} \underline{A} & \underline{B} \end{matrix} \\ \begin{bmatrix} .470 & .530 \\ .340 & .660 \end{bmatrix} \end{matrix}$$

$$P^7 = \begin{matrix} \\ \underline{A} \\ \underline{B} \end{matrix} \begin{matrix} \begin{matrix} \underline{A} & \underline{B} \end{matrix} \\ \begin{bmatrix} .391 & .609 \\ .390 & .610 \end{bmatrix} \end{matrix}$$

$$P^\ell = P^9 = P^k = \begin{matrix} \\ \underline{A} \\ \underline{B} \end{matrix} \begin{matrix} \begin{matrix} \underline{A} & \underline{B} \end{matrix} \\ \begin{bmatrix} .391 & .609 \\ .391 & .609 \end{bmatrix} \end{matrix}$$

Table 3

A Cyclic Chain and an Absorbing Chain For Transition Matrices Containing Two States

chain	Matrix
Cyclic	$P = \begin{array}{c} \\ \underline{A} \\ \underline{B} \end{array}\begin{array}{c} \begin{array}{cc} \underline{A} & \underline{B} \end{array} \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{array}$, $P^2 = \begin{array}{c} \\ \underline{A} \\ \underline{B} \end{array}\begin{array}{c} \begin{array}{cc} \underline{A} & \underline{B} \end{array} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array}$ $\vdots$ $\qquad\qquad$ $\vdots$ $P^{2k-1} = \begin{array}{c} \\ \underline{A} \\ \underline{B} \end{array}\begin{array}{c} \begin{array}{cc} \underline{A} & \underline{B} \end{array} \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{array}$, $P^{2k} = \begin{array}{c} \\ \underline{A} \\ \underline{B} \end{array}\begin{array}{c} \begin{array}{cc} \underline{A} & \underline{B} \end{array} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array}$
Absorbing	$P = \begin{array}{c} \\ \underline{A} \\ \underline{B} \end{array}\begin{array}{c} \begin{array}{cc} \underline{A} & \underline{B} \end{array} \\ \begin{bmatrix} .2 & .8 \\ 0 & 1 \end{bmatrix} \end{array}$, $P^k = \begin{array}{c} \\ \underline{A} \\ \underline{B} \end{array}\begin{array}{c} \begin{array}{cc} \underline{A} & \underline{B} \end{array} \\ \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \end{array}$

Table 4

Population Sizes of Transition Matrices
For Six Orders Of Markov Chain

m	Number of Transition Matrices	Percent of Absorbing Chains	Percent of Cyclic Chains	Percent of Reducible Chains	Percent of Regular Chains
2	1.0×10^{4}	1.98	0.01	----	98.01
3	1.4×10^{11}	1.01	0.00	0.00	98.99
4	9.8×10^{20}	1.00	0.00	0.00	99.00
5	2.0×10^{33}	1.00	0.00	0.00	99.00
6	7.9×10^{47}	1.00	0.00	0.00	99.00
7	3.7×10^{64}	1.00	0.00	0.00	99.00

[a] Based upon two digit transition probabilities.

Table 5

The Number Of Powers To Convergence For Percentile Ranks
For Six Orders Of Chain Using A Tolerance Of 0.01

	t					
P. R.	m[a] = 2	m = 3	m = 4	m = 5	m = 6	m = 7
1	0[b]	2	3	4	4	5
2	1	2	3	4	5	5
5	1	3	4	5	5	6
10	1	3	4	5	6	7
15	2	3	5	6	7	7
20	2	4	5	6	7	8
25	2	4	5	6	7	8
30	2	4	6	7	8	9
35	3	5	6	7	8	9
40	3	5	7	8	9	10
45	3	5	7	8	10	10
50	4	6	7	9	10	11
55	4	6	8	9	11	12
60	5	7	9	10	12	13
65	5	8	10	11	13	14
70	6	9	11	12	14	16
75	7	10	12	14	16	17
80	9	11	14	16	18	20
85	11	13	16	18	22	24
90	13	16	20	23	27	30
95	20	23	29	32	39	42
98	35	38	49	51	59	68
99	50	52	65	71	81	95

[a]The order of the transition matrices.

[b]One percent of the transition matrices were formed within the tolerance limit.

Table 6

Percentages and Cumulative Percentages Of Selected Numbers Of Powers To Convergence For Six Orders Of Chain Using A Tolerance Of 0.01

t[a]	m[b] = 2		m = 3		m = 4		m = 5		m = 6		m = 7	
	%	Cum %	%	Cum %	%	Cum %	%	Cum %	%	Cum %	%	Cum %
0	1.0	1.0										
1	13.3	14.3	0.3	0.3								
2	16.9	31.2	4.2	4.5	0.2	0.2	0.1	0.1				
3	13.8	45.0	12.7	17.3	3.8	4.0	0.7	0.8	0.1	0.1		
4	10.7	55.7	14.8	32.0	10.0	14.0	4.1	4.9	1.5	1.4	0.3	0.3
5	9.2	65.0	13.0	45.0	12.5	26.5	8.6	13.5	4.8	6.2	2.6	2.9
10		84.3		78.5		69.4		61.0		52.0		45.2
15		91.7		89.1		83.7		79.5		74.7		69.5
20		95.1		93.5		90.1		87.6		83.5		81.0
25		96.7		95.8		93.2		91.9		88.5		86.6
30		97.6		96.9		95.3		94.2		91.7		90.3
35		98.1		97.7		96.5		95.8		93.9		93.1
40		98.3		98.0		97.3		96.6		95.3		94.5
45		98.9		98.5		97.7		97.3		96.4		95.8
50		99.1		98.9		98.0		97.8		96.9		96.6
55		99.1		99.1		98.4		98.4		97.5		97.0
60		99.3		99.2		98.7		98.6		98.1		97.5

[a]The number of powers to convergence.

[b]The order of the transition matrices.

Table 7

Values and Percentages Of V_α For Six Orders Of Chain Using A Tolerance Of 0.01[a]

	Percentages					
V_α	m[b] = 2	m = 3	m = 4	m = 5	m = 6	m = 7
0.00	34.2	27.5	27.6	35.0	41.9	51.0
0.01	10.7	17.5	25.4	28.2	29.2	26.0
0.02	7.5	13.4	14.1	13.2	11.2	9.0
0.03	5.9	9.7	8.7	7.1	5.7	4.4
0.04	4.7	7.0	6.2	4.5	3.1	3.0
0.05	4.0	4.9	3.7	2.9	2.2	1.7
0.06	3.1	3.4	2.7	2.1	1.6	1.5
0.07	3.3	3.0	1.8	1.4	1.2	1.0
0.08	2.3	1.9	1.9	1.2	1.0	0.8
0.09	2.8	1.6	1.5	0.8	0.9	0.6
0.10	2.2	1.4	1.2	1.0	0.7	0.7
0.11	1.9	1.3	1.2	0.8	0.8	0.3
0.12	2.2	1.1	0.7	0.7	0.4	
0.13	1.8	1.4	0.8	0.6	0.1	
0.14	1.9	1.0	0.8	0.5		
0.15	1.2	0.7	0.6	0.0		
0.16	1.5	0.7	0.5			
0.17	1.4	0.8	0.4			
0.18	1.7	0.7	0.0			
0.19	1.5	0.5				
0.20	1.1	0.5				
0.21	1.2	0.1				
0.22	1.0					
0.23	0.9					
0.24	0.1					

[a]The variance of the elements in the limiting vector.

[b]The number of states in the transition matrices.

Table 8

Values and Cumulative Percentages Of V_α For Six Orders Of Chain Using A Tolerance Of 0.01[a]

	Cumulative Percentages					
V_α	m[b] = 2	m = 3	m = 4	m = 5	m = 6	m = 7
0.00	34.2	27.5	27.6	35.0	41.9	51.0
0.01	44.9	45.0	53.0	63.2	71.1	77.0
0.02	52.4	58.4	67.1	76.4	82.3	86.0
0.03	58.3	68.1	75.8	83.4	88.0	90.4
0.04	63.0	75.0	82.0	88.0	91.1	93.3
0.05	67.0	79.9	85.7	90.8	93.3	95.1
0.06	70.1	83.3	88.4	92.9	95.0	96.5
0.07	73.5	86.2	90.2	94.3	96.1	97.6
0.08	75.8	88.1	92.1	95.5	97.2	98.4
0.09	78.6	89.8	93.7	96.3	98.0	99.0
0.10	80.8	91.1	94.9	97.3	98.7	99.7
0.11	82.7	92.4	96.0	98.2	99.5	100.0
0.12	84.8	93.5	96.8	98.9	99.9	
0.13	86.6	94.9	97.5	99.5	100.0	
0.14	88.5	95.9	98.4	100.0		
0.15	89.7	96.6	99.0	100.0		
0.16	91.2	97.3	99.6			
0.17	92.6	98.1	100.0			
0.18	94.3	98.8	100.0			
0.19	95.7	99.3				
0.20	96.8	99.9				
0.21	98.0	100.0				
0.22	99.0					
0.23	99.9					
0.24	100.0					

[a]The variance of the elements in the limiting vector.

[b]The number of states in the transition matrices.

Table 9

Statistics Derived From Distributions of t and V_α For Six Orders Of Chain Using A Tolerance Of 0.01[a]

		m = 2	m = 3	m = 4	m = 5	m = 6	m = 7
t[b]	μ_1	6.9	9.1	12	13	15	17
	$\mu_2 x10^{-2}$	1.8	2.7	2.8	2.8	2.9	5
	$\mu_3 x10^{-4}$	4.2	9.4	5.8	4.9	3.9	13
	$\mu_4 x10^{-6}$	17	50	21	16	10	60
	$\beta_1 x10^{-1}$	29	45	15	11	6.1	13
	$\beta_2 x10^{-1}$	52	68	26	20	12	23
	κ_1	160	15	66	71	51	52
	κ_2	100	2600	65	34	15	69
V_α[b]	$\mu_1 x10^{2}$	5.0	3.4	2.6	1.9	1.5	1.2
	$\mu_2 x10^{3}$	3.9	1.9	1.2	0.71	0.49	0.36
	$\mu_3 x10^{4}$	3.1	1.5	0.81	0.44	0.27	0.18
	$\mu_4 x10^{5}$	5.4	2.2	0.96	0.44	0.23	0.14
	β_1	1.6	3.5	4.2	5.4	6.3	7.0
	β_2	3.5	6.2	7.1	8.6	9.7	10.5
	κ_1	-3.8	-4.1	-4.4	-5.0	-5.5	-6.0
	κ_2	- .50	-1.2	-1.5	-2.0	-2.3	-2.5
r_{t,V_α}[c]		-0.10	0.03	0.14	0.20	0.23	0.23
N[d]		4,901	4,949	4,933	4,958	4,945	4,949

[a] t is the number of powers to convergence, and V_α is the variance of the elements within the limiting vector.

[b] μ_1 is the first moment about the origin; μ_2, μ_3, μ_4 are the second, third and fourth moments about the mean respectively; β_1, β_2, and κ_1, are used to compute the criterion κ_2 of curve identification in the Pearson method of moments.

[c] The Pearson product moment correlation between t and V_α.

[d] The actual sample size of the distribution.

Table 10

Equations and Parameters For Curves Fitted To the Distributions Of V_{α} For Six Orders Of Chain Using A Tolerance Of 0.01[a]

m[b]	p[c]	q[c]	Distribution Function[d]
2	.558	10.6	$f(x) = \frac{1}{B(p,q)} x^{p-1} (1-x)^{q-1}$
3	.554	15.7	
4	.523	19.6	
5	.479	24.7	$0 \leq x \leq 1$
6	.437	28.7	$p, q \leq 0$
7	.383	31.5	

[a]The variance of the elements within the limiting vector.

[b]The number of states in the chain.

[c]p and q are parameters of the distribution functions.

[d]The distribution functions are Beta distributions.

Table 11

The Frequency Of Occurance Of Transition Matrices Of the Four Basic Types For Six Orders Of Chain Using A Tolerance Of 0.01

m[a]	Sample Size	Regular Chains	Absorbing Chains	Cyclic Chains	Reducible Chains
2	5000	4901	99	0	0
3	5000	4949	51	0	0
4	5000	4933	66	0	1
5	5000	4958	41	1	0
6	5000	4945	55	0	0
7	5000	4949	51	0	0

[a]The order of the transition matrices.